مقارنة بين الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل

د. أحمد إسماعيل الكحلوت *

الملخص : تهدف هذه الدراسة إلى مقارنة الخصائص السيكومترية لنمطين من الاختبارات الموضوعية الاختيار من متعدد والتكميل، وذلك من خلال فحص الفرضيات الآتية:

١) لا توجد فروق ذات دلالة إحصائية $(\alpha = ٠,٠٥)$ بين معاملات ثبات اختبار الاختيار من متعدد ومعاملات ثبات اختبار التكميل.

٢) لا توجد فروق ذات دلالة إحصائية $(\alpha = ٠,٠٥)$ بين تقديرات المحكمين لمدى ارتباط فقرات اختبار الاختيار من متعدد مع الاختبار نفسه، وبين تقديرات المحكمين لمدى ارتباط فقرات اختبار التكميل مع الاختبار نفسه.

٣) عدد العوامل المستخلصة من فقرات اختبار الاختيار من متعدد التي يزيد قيمة جذرها الكامن (Eigen value) عن واحد يساوي عدد العوامل المستخلصة من فقرات اختبار التكميل.

٤) لا يوجد فرق ذو دلالة إحصائية $(\alpha = ٠,٠٥)$ بين متوسط معاملات الصعوبة لاختبار الاختيار من متعدد ومتوسط معاملات الصعوبة اختبار التكميل.

٥) لا يوجد فرق ذو دلالة إحصائية $(\alpha = ٠,٠٥)$ بين متوسط معاملات التمييز لاختبار الاختيار من متعدد ومتوسط معاملات التمييز لاختبار التكميل.

وقد قام الباحث بإعداد اختبارين أحدهما من نوع الاختيار من متعدد والآخر من نوع التكميل بحيث تكون فقرات التكميل هي نفس فقرات اختبار الاختيار من متعدد بعد حذف البدائل، وتم تطبيق الاختبارين على عينة مكونة من ٤٥١ طالباً وطالبة من طلبة الصف السادس الأساسي تم اختيارهم عشوائيا من مجتمع الدراسة الذي يضم مدارس وكالة الغوث الدولية في الأردن.

وقد أظهرت النتائج أن معاملي ثبات اختبار التكميل المحسوبين بطريقة الإعادة وباستخدام معادلة كرونباخ ألفا أكبر من معاملي ثبات اختبار الاختيار من متعدد، وأظهرت نتائج اختبار (X^2) وجود اتفاق بين تقديرات المحكمين لمدى ارتباط فقرات كل اختبار مع الاختبار نفسه، كما بينت النتائج أن عدد العوامل ذات المعنى الجوهري المستخلصة من التحليل العاملي لاختبار التكميل يساوي عدد العوامل التي خُطط لقياسها، أي أربعة عوامل، في حين أن عدد العوامل المستخلصة من اختبار الاختيار من متعدد سبعة، بالإضافة إلى ما سبق فقد بينت النتائج أن متوسط معاملات الصعوبة لاختبار الاختيار من متعدد أكبر من متوسط معاملات الصعوبة لاختبار التكميل، وأن متوسط معاملات التمييز لاختبار التكميل أكبر من متوسط معاملات التمييز لاختبار الاختيار من متعدد .

* أستاذ مساعد بقسم علم النفس التربوي - كلية التربية - جامعة القدس المفتوحة .

خلفية الدراسة وأهميتها:

تعتمد دقة القرارات التي يتخذها المربون في جوانب متعددة من العملية التربوية على المعلومات التي تزودنا بها أدوات ووسائل القياس المتنوعة (عبدالله زيد الكيلاني و عبد الرحمن عدس، ١٩٩٤، ص٨). ولما كانت الاختبارات هي أكثر أدوات القياس استخداماً في المدرسة، فقد نالت عملية بنائها وتطويرها اهتماماً كبيراً من التربويين؛ وذلك لما يمكن أن تقدمه من بيانات موضوعية يمكن الاعتماد عليها في صنع القرارات السليمة التي تهدف إلى تحسين عمليتي التعليم والتعلم (عبد الله زيد الكيلاني و عبد الرحمن عدس، ١٩٩٤، ص ٥٨-٥٩)، إلا أن العديد من الدراسات وتقارير المؤتمرات والاجتماعات المتعددة للجان وجمعيات التقويم في العالم والعالم العربي خاصة تدل على وجود قصور في الاختبارات المدرسية (اليونسكو، ١٩٨٩ ص ٦-٨)، فالكثير منها يتصف بالغموض، وتعاني من نقص في تمثيل الأهداف، ومن ضعف في قدرتها التمييزية، أضف إلى ذلك الاستخدام العشوائي لأنماط الأسئلة الذي قد يكون بناء على قناعات تولدت لدى المعلمين بأفضلية نمط من الأسئلة على آخر دون توافر معلومات عن الخصائص السيكومترية لفقرات هذه الاختبارات (عبد الله زيد الكيلاني ١٩٩٤، ص ٩-١١، محمود الخطيب ص٧).

ونظراً للشكوى المستمرة من قصور الاختبارات فقد برزت الحاجة إلى تطويرها، وإجراء المزيد من البحوث والدراسات من أجل تحسينها ، إلا أن نتائج الدراسات العالمية والعربية قد بينت أن معظم هذه البحوث والدراسات يتجه نحو تطوير الجوانب الإدارية والتنظيمية التي تتناول في أغلب الأحيان قوانين ولوائح إجراءات تطبيقها، وشروط النجاح والرسوب أكثر مما يتجه نحو تطوير الجوانب المنهجية، والفنية، والتقنية في تصميم الاختبارات وتطويرها، التي تهدف إلى تحسين خصائصها السيكومترية من صدق وثبات، وخصائص الفقرات الاختبارية (اليونسكو، ١٩٨٩، ص ٧ -٩).

وعلى الرغم من أن صدق الاختبار وثباته يتأثران بالطريقة التي تصاغ بها الفقرات الاختبارية إلا أنه لا يزال كثير من التربويين لا يعتمدون معايير محددة في

اختيار نوع الفقرات الاختبارية (Herman 1992) ، إذ أن بعضهم يعدون اختبارات تتضمن فقرات مقالية معتقدين أنها الأكثر قدرة على قياس المستويات المتعددة من الأهداف، وأنها سهلة الإعداد، بينما يلاحظ أن فئة أخرى تعد اختبارات تتضمن فقرات موضوعية معتقدين أنها الأكثر شمولية وبالتالي الأكثر صدقاً وثباتاً من الاختبارات الأخرى، وهناك فريق ثالث يرى أنه من الأفضل أن يتضمن الاختبار فقرات موضوعية وأخرى مقالية بهدف الاستفادة من ميزات كل نوع، ولكن معظمهم ليس لديهم أدلة تجريبية على صدق ما يعتقدون.

وقد شاع في الآونة الأخيرة استخدام الاختبارات الموضوعية حيث يلاحظ أن المعلمين يستخدمون أنواعاً متعددة منها، فمنهم من يستخدم نوع واحد من هذه الاختبارات، ومنهم من يستخدم أكثر من نوع، ولكن اختبارات الاختيار من متعدد هي الأكثر استخداماً، وذلك لأن كثيراً من المعلمين والمقَيمين يرون أنها أفضل أنواع الاختبارات الموضوعية وأكثرها مرونة في قياس مستويات متعددة من الأهداف السلوكية، أي أنها لا تقتصر على قياس تذكر المعلومات بل تتعدى ذلك إلى قياس مستويات الفهم والاستيعاب والتطبيق والتحليل، وغيرها من النشاطات العقلية المركبة، كما يرون أيضاً أنها تتمتع بدلالات صدق وثبات كبيرتين، تفوق دلالات صدق وثبات الأنواع الأخرى من الاختبارات الموضوعية (عبد الله زيد الكيلاني، ١٩٩٤ ص٤٤).

وعلى الرغم من شيوع استخدام اختبارات الاختيار من متعدد لقياس تحصيل الطلبة في العديد من الأغراض التربوية(Gronlund 1988 p36) ، إلا أنها لا تزال تواجه نقداً متزايداً؛ فما يزال عامل التخمين من العوامل التي تهدد صدقها وثباتها؛ إذ لم يتمكن العاملون في مجال القياس والتقويم رغم محاولاتهم المتكررة من ضبط أثر التخمين باستخدام معادلات التصحيح التقليدية التي لا تستند إلى أسس نظرية علمية، أو بالاعتماد على طرق جديدة في التصحيح، مثل طريقة الاختيار الجزئي، وطريقة استبعاد البدائل غير الصحيحة (ساري سواقد، ١٩٩٢،ص١٣-١٤)، وفضلاً عن ذلك فإنه لا تتوفر أدلة تجريبية كافية تشير إلى أن هذا النوع من الاختبارات هو أكثر ثباتاً وصدقاً من جميع الاختبارات الموضوعية الأخرى، كما لا تتوفر أدلة تشير إلى أفضليتها في الكشف عن مدى

امتلاك الطلبة للمعرفة، وفي الكشف عن مدى قدرتهم على استخدامها، ويرى هيرمان أن الفهم الجديد لطبيعة تعلم الطلبة وسياقه يتطلب من الفاحصين التقليل من الاعتماد على الأسئلة الموضوعية من نوع الاختيار من متعدد، وزيادة الاهتمام بأسئلة الإجابة المنشأة (Herman 1992) ؛ وذلك لأن ما نختبره من خلال اختبارات الاختيار من متعدد لا يشكل دائماً تناظراً بين ما يقيسه الاختبار فعلاً وبين المعاني الحقيقية، أو عمليات التفكير التي نسعى لقياسها، وهذا قد يضعف صدق الاختبار وثباته، فإذا نظرنا إلى السؤال التالي الذي وضع لقياس مدى امتلاك الطلبة مهارة حل المعادلة من الدرجة الأولى، فقد يصل الطلبة إلى الإجابة الصحيحة باستخدام استراتيجية العكس، أي حل المعادلة من خلال التعويض بالبدائل الأربعة، معتمدين في ذلك على معرفتهم بمفهوم حل المعادلة وقدرتهم على التعويض بقيم عددية عن القيم الرمزية، وبناءاً على ذلك يعتقد المعلم أن هؤلاء الطلبة قد حققوا الهدف الذي خطط لقياسه، مع العلم بأنه من الممكن ألا يكونوا قد حققوا ذلك الهدف، أي أنهم قد لا يتقنون خطوات حل المعادلة التي تستخدم فيها قوانين التبديل، وقوانين التجميع، وخاصية العنصر المحايد لعملية الجمع، والعنصر المحايد لعملية الضرب.

السؤال : حل المعادلة ٣(س+٥) = ٢٥-٢س هو

أ) ٤ ب) ٦ ج) ٢ د) ٨.

أما الأسئلة ذات الإجابة المنشأة فإنها لا تسمح باستخدام استراتيجية الرجوع إلى الخلف ، في حين كما لاحظنا في المثال السابق أن أسئلة الاختيار من متعدد تسمح بذلك (Bridgeman 1992 pp 255-256) .

كما أن الأسئلة ذات الإجابة المنشأة تقلل من تأثير التغذية الراجعة غير المقصودة، بمعنى أنه إذا كانت الإجابة التي فكر فيها المفحوص ليست من بين البدائل المعطاة؛ فإنه يعرف بوجود خطأ معين، ويحاول تطبيق استراتيجية بديلة لمعرفة الجواب الصحيح (Herman 1992) .

ويقول باندي إن الأسئلة ذات الإجابة المنشأة تتمتع بمزايا متعددة يذكر منها المزايا التالية:

- تمنح الطلبة فرصة ليعبروا عن أفكارهم.
- تقيس جميع مستويات الأهداف التي تقيسها أسئلة الاختيار من متعدد.

- تقلل أخطاء القياس بحذف التخمين العشوائي.
- تمنح الطلبة فرصة لبناء إجابتهم بدلاً من اختيار إجابة واحدة من بين عدة إجابات معطاة.
- تسمح للطلبة بعرض مدى فهمهم للمشكلة (Pandy, 1990, p3).

مما سبق نستطيع القول أن الأسئلة المفتوحة ذات الإجابة المنشأة تتمتع بصفات تجعلها أكثر قدرة إلى حد ما على قياس ما يمتلكه الطلبة من معارف ومهارات متعددة من أسئلة الاختيار من متعدد، إلا أن هذا لا يؤكد لنا أن الاختبارات التي تتضمن أسئلة ذات الإجابة المنشأة تمتع بخصائص سيكومترية أفضل من الخصائص السيكومترية للاختبارات التي تتضمن أسئلة الاختيار من متعدد.

وقد قام الباحث بمسح للدراسات التي أجريت بهدف المقارنة بين اختبارات الاختيار من متعدد واختبارات الإجابة المنشأة، واختبارات التكميل باعتبارها حالة خاصة من اختبارات الإجابة المنشأة، حيث وقف الباحث على عدد قليل من هذه الدراسات نتائجها ليست متسقة.

فقد توصل (بينيت وآخرون) من خلال التحليل العاملي لفقرات الاختيار من متعدد، وفقرات اختبارات الإجابة المنشأة إلى أن نمطي الاختباراين يتمتعان بدرجة عالية من صدق البناء وأنه يمكن بناء اختبارين لهما العدد نفسه من العوامل التي تفسر التباين في البيانات المأخوذة من كليهما، ومن الجدير ذكره أن اختبارات الإجابة المنشأة واختبارات الاختيار من متعدد التي استخدمت في هذه الدراسة لم تكن متكافئة في المحتوى (Bennett et, al., 1990, pp.85-88).

وقد أظهرت نتائج دراسة ثايسن و آخرين التي أستخدم فيها اختباران غير متساويين في عدد الفقرات الاختبارية أحدهما في علوم الحاسوب والثاني في الكيمياء، أن عدد العوامل المستخلصة من فقرات الاختيار من متعدد أكبر من عدد العوامل المستخلصة من الفقرات ذات الإجابة المنشأة، ولكن التحليل المنطقي المستند إلى أراء محكمين بين أن فقرات الاختيار من متعدد تقيس ما تقيسه فقرات الإجابة المنشأة (Thissen et, al.,1994, 117-120).

وقد قارنت دراسة بردجمان بين نتائج الطلبة على اختبار الاختيار من متعدد مع نتائجهم على اختبار الإجابة المنشأة، وتبين أن متوسط علامات الطلبة في اختبار الإجابة المنشأة أقل من متوسط علاماتهم في اختبار الاختيار من متعدد، على الرغم من أن ترتيب الطلبة على الاختبارين متقارب بدرجة كبيرة، كما بينت النتائج أن ثبات اختبارات الإجابة المنشأة أعلى من ثبات اختبارات الاختيار من متعدد . (Bridge man 1994, pp 265-268) .

أما دراسة كدالا فقد بينت النتائج وجود فروق دالة إحصائياً بين متوسط علامات الطلبة على اختبار الاختيار من متعدد الذي يقيس مهارة إجراء العمليات الحسابية واختبار الإجابة المنشأة الذي يقيس المهارة نفسها، وذلك في الصفين الثاني والثالث الأساسيين بينما لم تكن الفروق دالة إحصائياً بين نتائج طلبة الصفوف من الرابع إلى السادس (Gadalla, 1999) .

كما بينت نتائج دراسة بادجر التي استخدم فيها اختبار في العلوم وآخر في الرياضيات، يتضمن كل منهما فقرات من نوع الاختيار من متعدد يزيد عددها عشرين ضعفاً عن عدد الفقرات التي تتطلب إجابة منشأة، أن أداء طلبة الصف الثاني عشر على فقرات الاختيار من متعدد كان أعلى من أدائهم على فقرات الإجابة المنشأة (Badger, 1990) .

ويقول بولاك في ورقة قدمها ضمن سلسلة من حلقات البحث والمناقشة في مركز الاختبارات التربوية (ETS) « أنه على الرغم من أن عملية تمثيل المحتوى من خلال الأسئلة ذات الإجابة المنشأة لا يكون شاملاً بدرجة كافية، إلا أنها أكثر دقة في قياس السمة المراد قياسها من أسئلة الاختيار من متعدد، أي أكثر صدقاً، كما أنها توفر فرصة أكبر لقياس ما يمكن أن يفعله الطالب أكثر من قياس ما يعرفه » ، ولكنه يؤكد أن اختبارات الاختيار من متعدد أكثر ثباتاً من الاختبارات الأخرى ، ولهذا فإنه يقترح الدمج بين النمطين لتحقيق صدق وثبات أعلى (Pollack,1990) .

يتضح مما سبق أن الدراسات السابقة التي قارنت بين نمطين من الاختبارات الاختيار من متعدد والاختبارات ذات الإجابة المنشأة ولم تتوصل إلى نتائج متسقة، كما أنها لم تستخدم اختبارات متساوية في عدد الفقرات، بالإضافة إلى أن محتوي الفقرات الاختبارية مختلف؛ وعليه فإن الحاجة ماسة إلى مزيد من الدراسات تتناول

الخصائص السيكومترية لاختبارات تقيس نفس المحتوى، ولها العدد نفسه من الفقرات، وقد رأى الباحث أن يجري مقارنة بين اختبارات الاختيار من متعدد واختبارات التكميل التي تعد حالة خاصة من اختبارات الإجابة المنشأة.

مشكلة الدراسة :

سعت هذه الدراسة إلى مقارنة الخصائص السيكومترية من صدق وثبات وتحليل للفقرات الاختبارية لنمطين من الفقرات الاختبارية الموضوعية الاختيار من متعدد والتكميل. وبالتحديد فقد حاولت الدراسة الإجابة عن السؤال الآتي:

- هل تختلف الخصائص السيكومترية لاختبارات التكميل عن الخصائص السيكومترية لاختبارات الاختيار من متعدد؟ وللإجابة عن هذا السؤال فعلينا أن نجيب عن الأسئلة الآتية:

١) هل تختلف معاملات ثبات اختبار الاختيار من متعدد عن معاملات ثبات اختبار التكميل؟

٢) هل تختلف تقديرات المحكمين لمدى ارتباط فقرات اختبار الاختيار من متعدد مع الاختبار نفسه، عن تقديراتهم لمدى ارتباط فقرات اختبار التكميل مع الاختبار نفسه؟

٣) هل يختلف عدد العوامل الجوهرية التي يزيد قيمة الجذر الكامن لها عن واحد . المستخلصة من التحليل العاملي لفقرات اختبار الاختيار من (Eigen value>1) متعدد عن عدد العوامل الجوهرية المستخلصة من اختبار التكميل؟

٤) هل تختلف معاملات الصعوبة لفقرات اختبار الاختيار من متعدد عن معاملات الصعوبة لفقرات اختبار التكميل؟

٥) هل تختلف معاملات التمييز لفقرات اختبار الاختيار من متعدد عن معاملات التمييز لفقرات اختبار التكميل؟

للإجابة عن الأسئلة السابقة فقد اقترحت الفرضيات الآتية:

١) لا توجد فروق ذات دلالة إحصائية (α < ٠,٠٥) بين معاملات ثبات اختبار الاختيار من متعدد ومعاملات ثبات اختبار التكميل.

٢) لا توجد فروق ذات دلالة إحصائية (α < ٠,٠٥) بين تقديرات المحكمين لمدى ارتباط فقرات اختبار الاختيار من متعدد مع الاختبار نفسه، وبين تقديرات المحكمين لمدى ارتباط فقرات اختبار التكميل مع الاختبار نفسه.

٣) عدد العوامل المستخلصة من فقرات اختبار الاختيار من متعدد التي يزيد قيمة جذرها الكامن عن واحد (Eigen value>1) يساوي عدد العوامل المستخلصة من فقرات اختبار التكميل .

٤) لا يوجد فرق ذو دلالة إحصائية(α < ٠,٠٥) بين متوسط معاملات الصعوبة لاختبار الاختيار من متعدد ومتوسط معاملات الصعوبة لاختبار التكميل.

٥) لا يوجد فرق ذو دلالة إحصائية (α < ٠,٠٥) بين متوسط معاملات التمييز لاختبار الاختيار من متعدد ومتوسط معاملات التمييز لاختبار التكميل.

تعريف مصطلحات الدراسة:

الخصائص السيكومترية : أي دلالات الصدق للاختبار ودلالات الثبات بالإضافة إلى خصائص الفقرات الاختبارية، التي تتضمن معاملات الصعوبة ومعاملات تمييز الفقرات الاختبارية.

- **معامل الصعوبة :** النسبة بين عدد الطلبة الذين أجابوا على الفقرة الاختبارية إجابة صحيحة وعدد الطلبة الذين حاولوا الإجابة عليها
- **معامل التمييز :** مؤشر على قدرة الفقرة على التمييز بين الفئة العليا من الطلبة والفئة الدنيا من الطلبة، ويمكن تقديره من خلال إيجاد معامل ارتباط علامات الطلبة على الفقرة مع علاماتهم على الاختبار.
- **الجذر الكامن** (Eigen value) : النسبة بين التباين في المتغير التابع الذي يفسره عامل من العوامل وبين التباين المتبقي (التباين الخطأ) .
- **اختبارات الإجابة المنشأة :** أي الاختبارات التي تتطلب من المفحوص أن يقوم بإنشاء الإجابة وليس اختيارها من بين عدد من الإجابات، ويعد اختبار التكميل حالة خاصة منها.

مجتمع الدراسة والعينة :

يتشكل مجتمع الدراسة من جميع طلبة الصف السادس الابتدائي في مدارس

وكالة الغوث الدولية في الأردن في العام الدراسي ١٩٩٩/ ٢٠٠٠ البالغ عددهم ١٣٣٤٦ منهم ٦٧٧٩ من الإناث و٦٥٦٧ من الذكور موزعين في ٨٨ مدرسة للإناث و٨٥ مدرسة للذكور. أما عينة الدراسة فإنها تكونت من ٤٥١ طالباً وطالبة من طلبة الصف السادس الأساسي موزعين في ١٢ شعبة، وتعد هذه العينة ممثلة للمجتمع تمثيلاً جيداً؛ لأن عدد أفرادها يزيد عن خمسة أضعاف عدد الفقرات الاختبارية، وهو العدد المقترح لدراسة الخصائص السيكومترية للاختبارات (خليل عليان، ١٩٩٤ ص١٥٥)،.

وفيما يلي إجراءات اختيار العينة:

١ - حصر مدارس وكالة الغوث في الأردن التي تضم الصف السادس.

٢ - اختيار ٦ مدارس من مدارس الإناث بطريقة عشوائية، ومثلها من مدارس الذكور.

٣ - اختيار شعبة واحدة من كل مدرسة، وذلك بطريقة عشوائية.

أدوات الدراسة :

قام الباحث بإعداد اختبارين يتضمن كل منهما ٢٣ فقرة اختبارية جميعها على شكل رسوم لأشكال هندسية تمثل إحدى العمليات الرياضية الأربع المعرفة على الكسور العادية، بحيث يتضمن الاختبار الأول فقرات اختبارية من نوع الاختيار من متعدد، والثاني يتضمن فقرات من نوع التكميل. وقد تم إعداد فقرات اختبار الاختيار من متعدد على شكل عبارة ناقصة يتم إكمالها من إحدى البدائل الأربع التي تتبع هذه العبارة الناقصة، أما فقرات اختبار التكميل فقد تم إعدادها بحذف بدائل فقرات الاختيار من متعدد، ولذلك لم تكن هناك حاجة لأن يرفق هذا الاختبار. ومن الجدير ذكره أنه قد تمت الاستفادة في عملية إعداد الفقرات الاختبارية من منهاج الرياضيات لصفوف الرابع والخامس والسادس في الأردن، وفيما يلي الخطوات التي أتبعت في إعداد فقرات اختبار الاختيار من متعدد.

١ - أعد الباحث أربع مجموعات من الفقرات الاختبارية تقيس كل منها مفهوماً واحداً من مفاهيم العمليات الرياضية الأربع على الكسور (الجمع والطرح

والضرب والقسمة) بحيث تتضمن كل مجموعة ١٠ فقرات اختبارية من نوع الاختيار من متعدد لكل فقرة أربعة بدائل.

٢ - عرضت الفقرات الاختبارية على سبعة من المختصين المتميزين في تدريس الرياضيات، أربعة من المعلمين وثلاثة من المشرفين التربويين، حيث قام كل منهم بوضع ملاحظاته عن كل فقرة اختبارية وفقاً للمعايير الآتية:

١) وضوح اللغة .

٢) سلامة اللغة العلمية .

٣) الدقة في صياغة الفقرات الاختبارية.

٤) قدرة الفقرة الاختبارية على قياس فهم الطلبة للمفهوم الذي وضعت لقياسه.

٥) فاعلية البدائل من وجهة نظر منطقية.

٦) الوزن النسبي لكل فقرة اختبارية.

٣ - قام الباحث بحذف الفقرات التي قرر خمسة من المختصين أنها غير مرتبطة بالاختبار ارتباطاً قوياً، وكذلك حذف الفقرات التي تتضمن بدائل غير فاعلة، ولأغراض سهولة التصحيح وتحليل النتائج قام الباحث بحذف الفقرات التي قرر خمسة من المختصين أن وزنها النسبي يزيد عن واحد، أما الفقرات التي لم تحقق الشروط الثلاثة الأولى فقد تم تعديلها وفق الملاحظات التي أبداها المختصون.

٤ - تم تجريب الفقرات المتبقية وعددها ٣١ فقرة على عينة صغيرة من مجتمع الدراسة مكونة من ٢٠ طالباً تم اختيارهم من مجتمع الدراسة بطريقة عشوائية، وذلك بهدف تحديد الزمن المناسب لتطبيق الاختبار، ودراسة مدى وضوح الفقرات الاختبارية للطلبة، ومدى فاعلية البدائل تجريبياً.

٥ - حسب معدل الزمن المناسب للإجابة عن الفقرة الاختبارية الواحدة، حيث لوحظ أنه يعادل تقريباً دقيقة ونصف الدقيقة، وبناءً على ذلك تم تقدير مدة الاختبار بحوالي ٤٠ دقيقة، ولضبط عامل الزمن فقد خصص مدة زمنية لاختبار التكميل مساوية للمدة الزمنية لاختبار الاختيار من متعدد.

٦ - حذف الفقرات التي قرر ٨٠٪ من الطلبة عدم وضوحها، كما حذفت الفقرات التي كانت تتضمن بدائل غير فاعلة، أي البدائل التي لم يختارها أحد من الطلبة، وقد بلغ عدد الفقرات المتبقية ٢٣ فقرة اختبارية .

٧ - إجراءات التطبيق .

بعد أن أصبح الاختباران في صورتهما النهائية فقد تم دراسة خصائصهما السيكومترية وفق الإجراءات الآتية:

١ - أعطيت جميع الفقرات العلامة (١) إذا كانت الإجابة صحيحة والعلامة (٠) إذا كانت الإجابة خطأ.

٢ - عرضت فقرات كل اختبار من الاختبارين على أربعة من المحكمين المختصين في الرياضيات وفي القياس والتقويم، اثنان منهم من حملة الدكتوراه، واثنان من حملة الماجستير، وطلب منهم وضع تقديراتهم لمدى ارتباط كل فقرة من فقرات الاختبارين مع الاختبار الذي تنتمي إليه، وذلك باختيار أحد الأعداد من صفر إلى أربعة، حيث حددت القيمة (٤) لتدل على وجود ارتباط كبير جداً بين الفقرة والاختبار، والقيمة (٠) تدل على عدم وجود ارتباط، ثم حسبت معدلات تقديراتهم لكل فقرة في كل اختبار.

٣ - تم تطبيق كل اختبار مرتين بينهما أسبوعان على عينة مكونة من(١٠٠) طالب وطالبة ليست من عينة الدراسة، وذلك من أجل تقدير معامل الاستقرار.

٤ - تم تطبيق كلا الاختبارين على عينة الدراسة المكونة من ٤٥١ طالباً وطالبة حيث طبق اختبار التكميل أولاً، ثم طبق اختبار الاختيار من متعدد بعد أسبوع على العينة نفسها، وذلك لتلافي احتمال استفادة الطالب من البدائل إذا ما طبق اختبار الاختيار من متعدد قبل اختبار التكميل.

تصميم الدراسة والتحليل الإحصائي :

طبق الباحث الدراسة على مجموعة واحدة عوضاً عن مجموعتين متكافئتين، وذلك من أجل تقليل الأخطاء بين المجموعات، وتقليل أخطاء المعاينة التي تنشأ من تجزئة المجموعة إلى نصفين. أما التحليلات الإحصائية التي استخدمت في فحص فرضيات الدراسة فقد جرى حسابها باستخدام برمجية SPSS حيث استخدم معامل ارتباط بيرسون لتقدير معامل الاستقرار، ومعامل كرونباخ ألفا لتقدير الاتساق الداخلي، كما أُستخدم الإحصائي (X^2) لفحص فرضية التوافق بين تقديرات المحكمين، ولفحص فرضية العوامل أُستخدم التحليل العاملي بطريقة المكونات

الرئيسية (Principal Components) حيث اعتبرت العوامل التي يزيد قيمة جذرها الكامن عن واحد عوامل جوهرية، واستخدم اختبار (T) لفحص فرضيتي الصعوبة والتمييز، وقد حسب معامل الصعوبة من خلال إيجاد نسبة عدد الطلبة الذين أجابوا على الفقرة الاختبارية إجابة صحيحة إلى عدد الطلبة الذين حاولوا الإجابة عن الفقرة، أما معامل التمييز للفقرة فقد تم تقديره من خلال حساب معامل الارتباط بين علامات الطلبة على الفقرة وبين علاماتهم على الاختبار كله باستثناء علاماتهم على تلك الفقرة.

النتائج :

سعت الدراسة لفحص خمس فرضيات صفرية، حيث رُفضت جميع هذه الفرضيات عدا الفرضية الثانية. وفيما يلي توضيح للنتائج التي توصلت إليها الدراسة:

أولاً : النتائج المرتبطة بالفرضية الأولى : أظهرت النتائج المبينة في الجدول (١) أن كلا الاختبارين يتمتع بدلالات ثبات جيدة، كما نستنتج من الجدول أن معاملات ثبات اختبار التكميل المقدرة بطريقة الإعادة (معامل الاستقرار)، وكذلك معاملات الثبات المقدرة باستخدام معادلة كرونباخ ألفا أعلى من معاملات ثبات اختبار الاختيار من متعدد المقدرة بالطريقتين السابقتين.

جدول (١) : معاملات الثبات لاختباري الاختيار من متعدد والتكميل .

طريقة تقدير الثبات	الاختبار من متعدد		التكميل	
	معامل الثبات	مستوى الدلالة	معامل الثبات	مستوى الدلالة
طريقة الإعادة	٠,٧٥	٠,٠٠١	,٨١	,٠٠٠١
كرونباخ ألفا	٠,٧٤	٠,٠٠١	,٩٠	,٠٠٠١

كما تبين النتائج الموضحة في جدول (٢) أن الفروق بين معاملات الثبات هي فروق دالة إحصائياً على مستوى الدلالة ($\alpha < ٠,٠١$) حيث أستخدم اختبار Z بعد أن تم تحويل قيم معاملات الثبات إلى قيم معيارية Z باستخدام العلاقات الآتية:

$$Z_1 = \frac{1}{2}\ln\left(\frac{1+|r_1|}{1-|r_1|}\right) \qquad \text{حيث} \qquad Z = \frac{z_1 - z_2}{\sigma_{z_1 - z_2}}$$

$$\sigma_{z_1 - z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad \text{(Glass \& Hopkins, 1984 pp 305-306)}$$

جدول (٢) : قيم Z ومستويات الدلالة للفروق بين معاملات الثبات

طريقة حساب الثبات	قيم Z	مستوى الدلالة
طريقة الإعادة	٢,٣٤	٠,٠١
كرونباخ ألفا	٧,٩	٠,٠٠٠١

ثانياً النتائج المرتبطة بالفرضية الثانية : يبين الجدول (٣) معدلات تقديرات المحكمين لمدى ارتباط كل فقرة اختبارية مع الاختبار الذي تنتمي إليه، حيث يظهر من الجدول أن معدلات تقديرات المحكمين لمدى ارتباط غالبية فقرات كل من اختباري التكميل والاختيار مع الاختبار نفسه تساوي ٣ درجات أو أكثر، أي أن معدلات التقدير لمعظم الفقرات تساوي ٧٥٪ أو أكثر، باستثناء فقرتين من فقرات اختبار الاختيار من متعدد وثلاث فقرات من فقرات اختبار التكميل حيث يزيد معدل تقديرات المحكمين لهذه الفقرات عن ٦٢٪ ، وتعد هذه النتيجة دليلاً يدعم الصدق المنطقي للاختبارين، أي أن كلا الاختبارين يتمتع بدلالات صدق جيدة .

جدول (٣) : معدلات تقديرات المحكمين لمدى ارتباط كل فقرة للاختبار الذي تنتمي إليه

الاختيار من متعدد				التكميل			
الرقم	المعدل	الرقم	المعدل	الرقم	المعدل	الرقم	المعدل
١	٣,٢٥	١٣	٣,٥	١	٣,٢٥	١٣	٣,٥
٢	٣,٢٥	١٤	٣,٢٥	٢	٣	١٤	٣
٣	٣,٥	١٥	٣,٥	٣	٣,٥	١٥	٣,٧٥
٤	٣,٧٥	١٦	٣	٤	٣,٥	١٦	٣
٥	٣,٢٥	١٧	٢,٧٥	٥	٣,٢٥	١٧	٢,٧٥
٦	٣	١٨	٣,٥	٦	٣,٢٥	١٨	٢,٥
٧	٣,٢٥	١٩	٣,٢٥	٧	٣,٧٥	١٩	٣,٢٥
٨	٣,٥	٢٠	٣,٧٥	٨	٣,٥	٢٠	٣,٥
٩	٣	٢١	٣,٢٥	٩	٣,٥	٢١	٣,٢٥
١٠	٢,٥	٢٢	٣,٥	١٠	٣,٢٥	٢٢	٣,٥
١١	٣	٢٣	٣	١١	٢,٧٥	٢٣	٣,٢٥
١٢	٣,٢٥			١٢	٣,٢٥		

كما بينت نتائج اختبار (X^2) وجود توافق بين تقديرات المحكمين لمدى ارتباط فقرات اختبار الاختيار من متعدد مع الاختبار نفسه مع تقديراتهم لمدى ارتباط فقرات اختبار التكميل مع الاختبار نفسه، حيث بلغت قيمة (X^2 ٣٣,٥٠٤) بدرجات حرية ٢٥ ومستوى دلالة ٠,١١٨٩ ، وهذه النتيجة تجعلنا نميل إلى قبول الفرضية الصفرية، التي تفترض وجود توافق بين تقديرات المحكمين، وبذلك نستطيع القول أن الاختبارين لهما دلالات صدق منطقي متكافئة.

ثالثاً النتائج المرتبطة بالفرضية الثالثة : يوضح الجدول (٤) عدد العوامل المفسرة للتباين ذات المعنى الجوهري، أي التي تزيد قيم جذورها الكامنة عن واحد ليست متساوية في الاختبارين.

الجدول (٤) :العوامل المستخلصة من التحليل العاملي لفقرات اختبار التكميل واختبار الاختيار من متعدد وقيم الجذور الكامنة ونسبة التباين المفسرة والنسبة التراكمية للتباين المفسرة

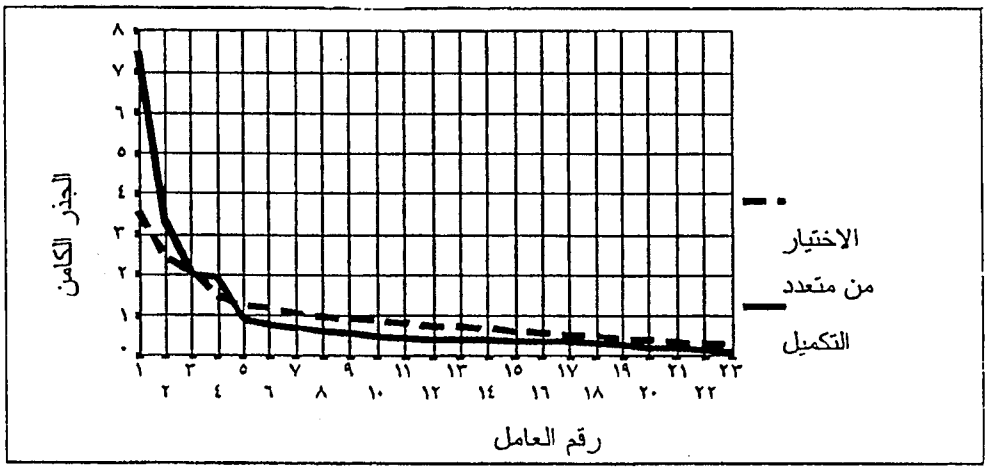
العامل	اختبار التكميل			اختبار الاختيار من متعدد		
	الجذر الكامن	نسبة التباين		الجذر الكامن	نسبة التباين	
		المفسرة	التراكمية		المفسرة	التراكمية
١	٧,٤٧٦	٣٢,٥	٣٢,٥	٣,٥٥٨	١٥,٥	١٥,٥
٢	٣,٣٦١	١٤,٦	٤٧,١	٢,٤٣٦	١٠,٦	٢٦,١
٣	٢,٠١٢	٨,٧	٥٥,٩	٢,٠٧١	٩,٠	٣٥,١
٤	١,٩٢٦	٨,٤	٦٤,٢	١,٤٩٨	٦,٥	٤١,٦
٥	٠,٩٠٦	٣,٩	٦٨,١	١,٢٣٥	٥,٤	٤٧
٦	٠,٧٧١	٣,٤	٧١,٥	١,١٨٣١	٥,١	٥٢,١
٧	٠,٧٢١	٣,١	٧٤,٧	١,٠٧٧	٤,٧	٥٦,٨
٨	٠,٦٠٨	٢,٦	٧٧,٣	٠,٩٧٦	٤,٢	٦١
٩	٠,٥٦٦	٢,١	٧٩,٨	٠,٩٥٧	٤,٢	٦٥,٢
١٠	٠,٤٨٨	٢,٥	٨١,٩	٠,٨٩٣	٣,٩	٦٩,١
١١	٠,٤٤٨	١,٩	٨٣,٨	٠,٨١٣	٣,٥	٧٢,٦
١٢	٠,٤٢٦	١,٩	٨٥,٧	٠,٧٥٢	٣,٢	٧٥,٩
١٣	٠,٤١٨	١,٨	٨٧,٥	٠,٧٣١	٣,٢	٧٩,١
١٤	٠,٣٩٧	١,٧	٨٩,٢	٠,٧٠٢	٣	٨٢,١
١٥	٠,٣٩١	١,٧	٩١	٠,٦٢٠	٢,٧	٨٤,٨
١٦	٠,٣٦٩	١,٦	٩٢,٦	٠,٥٩٠	٢,٦	٨٧,٤
١٧	٠,٣٥٤	١,٥	٩٤,١	٠,٥٣٥	٢,٣	٨٩,٧
١٨	٠,٣٤٤	١,٥	٩٥,٦	٠,٥١٨	٢,٣	٩٢,٠
١٩	٠,٢٨٨	١,٣	٩٦,٩	٠,٤٣٩	١,٩	٩٣,٩
٢٠	٠,٢٢٥	١,١	٩٧,٨	٠,٤٠٦	١,٧	٩٥,٦
٢١	٠,١٩٨	٠,٩	٩٨,٧	٠,٣٦٠	١,٦	٩٧,٢
٢٢	٠,١٦٦	٠,٧	٩٩,٤	٠,٣٢٩	١,٤	٩٨,٦
٢٣	٠,١٣٣	,٦	١٠٠	٠,٣١٢	١,٤	١٠٠

ويتبين من الجدول (٤) أن عدد العوامل المستخلصة من التحليل العاملي لاختبار التكميل يساوي أربعة عوامل تفسر ٦٤,٢٪ من التباين، في حين أن عدد العوامل المستخلصة من اختبار الاختيار من متعدد سبعة عوامل تفسر فقط ٥٦,٨٪ من التباين، الأمر الذي يجعلنا نميل إلى رفض الفرضية الصفرية وقبول الفرضية

البديلة، وحيث أن نتيجة التحليل العاملي لاختبار التكميل متفقة مع التحليل المنطقي لفقرات الاختبار، إذ بني الاختبار ليقيس أربع عمليات حسابية، بينما لم تتفق نتائج التحليل العاملي لاختبار الاختيار من متعدد مع التحليل المنطقي لفقرات الاختبار، نستنج من ذلك أن الصدق العاملي لاختبار التكميل أفضل من الصدق العاملي لاختبار الاختيار من متعدد.

كما أن الشكل (١) يوضح النتائج التي توصلنا إليها في الجدول (٤) .

الشكل (١) العوامل المستخلصة من التحليل العاملي لاختبار التكميل واختبار الاختيار من متعدد وقيم الجذور الكامنة



ولمعرفة فيما إذا كان التحليل العاملي يعمل على تجميع الفقرات الاختبارية على العوامل ذات المعنى الجوهري بنفس الطريقة التي تتجمع فيها الفقرات على العوامل كما خطط الباحث، فقد قام الباحث بإجراء التدوير المتعامد للعوامل الناتجة من التحليل العاملي لكل اختبار بطريقة (Varimax) ، وحسبت تشبعات الفقرات على كل عامل من هذه العوامل. ويبين الجدول (٥) تشبعات فقرات اختبار التكميل على العوامل الأربعة المستخلصة من هذا الاختبار.

جدول (٥) : تشبعات الفقرات بكل عامل من العوامل الأربعة المستخلصة من اختبار التكميل

أرقام الفقرات	العوامل المستخلصة			
	الأول	الثاني	الثالث	الرابع
١٨	٠,٧٨٢	٠,٠٤٤٧_	٠,١٧٠٨	٠,٠٦٤
٣	٠,٧٦٢٤	٠,١٤٩٧	٠,٢١١٧	٠,٠٦٩٢
١٩	٠,٧٥٧٧	٠,١١٩٤	٠,٠٦٩	٠,٠٣٢٦
١	٠,٧٤٨٢	٠,٠٤٧٨	٠,٢٥٣٢	٠,٠٧٨٦
١٣	٠,٧٤٥٣	٠,٠١٨٧	٠,٢٢٣٧	٠,٠٦١٥
٤	٠,٧١٣٤	٠,٠٩٨١	٠,١٣٥٧٥	٠,١٩٣٢
١١	٠,٦٨٢٢	٠,١٨٥١	٠,٢٨٩٣	٠,١٦٣٢
١٦	٠,٠٥٧١	٠,٨٧٥٦	٠,١٠٠١	٠,٠٩٠٨
١٠	٠,٠٨٦٥	٠,٨٣٦١	٠,١٧٤٨	٠,٠٩٠٨
١٥	٠,١٤٧٥	٠,٨٢٤٨	٠,١٧٠٨	٠,٠١٩٥
٧	٠,١٥٩١	٠,٨١٠١	٠,٠٤٢٨	٠,١٦٦٢
٦	٠,٠٣٦٧٥-	٠,٧٧٩٦	٠,١١٤٠	٠,٠٨٥١
٩	٠,٠٨٨٣	٠,٧٦٢٠	٠,٠٤٣٣	٠,٠٢٥٤_
١٢	٠,١٤٣٠	٠,١٣٧٦	٠,٨١٤٥	٠,٠٨٤٦
٨	٠,١٥٥٧	٠,٠١٧٩	٠,٧٦٧١	٠,١١٤٣
١٤	٠,٢٣٠٠	٠,٠٦٣٤	٠,٧٦١٩	٠,١٦٧٩
٥	٠,٢٢٥٤	٠,١٩٢٣٣	٠,٧٦١٩	٠,٠١٢٥_
٢	٠,٢٠٦٩	٠,١٤٩٣	٠,٧٥٠٩	٠,١٢٦٣
١٧	٠,٣٩٧٤٧	٠,١٥٥٤	٠,٦٥٦٣	٠,١١٤٩
٢١	٠,٠٥١٥	٠,٠٠٤٥	٠,١٠٣٨	٠,٧٧٧٥
٢٢	٠,٠٦٦٨	٠,٠٦٨٤	٠,١٢٧٠	٠,٧٦٧٨
٢٣	٠,١٠٧٢	٠,١٠٢٢	٠,٠٢٩٧	٠,٧٣٠٦
٢٠	٠,٢٠٨٦	٠,١٣٣٠	٠,٠٩٢١	٠,٦٨٨٢

يتضح من الجدول أن الفقرات الاختبارية التي تقيس مفهوم الجمع قد تجمعت حول العامل الأول، وأن فقرات الطرح تجمعت حول العامل الثاني، وان فقرات الضرب قد تجمعت حول العامل الثالث، وأن فقرات القسمة قد تجمعت حول العامل الرابع وهذه النتيجة تنسجم مع النتيجة التي توصلنا إليها من التحليل المنطقي في بند أدوات الدراسة الفقرة السادسة، وكلاهما يتفق مع الطريقة التي وزع بها الباحث الفقرات في عملية الإعداد الأولى، وتعد هذه النتيجة دلالة صدق جيدة لاختبار التكميل.

جدول (٦) : تشبعات الفقرات بكل عامل من العوامل الأربعة المستخلصة من اختبار التكميل

أرقام الفقرات	العوامل المستخلصة						
	الأول	الثاني	الثالث	الرابع	الخامس	السادس	السابع
١٣	٠,٨٢١٥	٠,٠٤٤٧	٠,٠٠٩_	٠,٠٥٣٣	٠,٠٤٧_	٠,٠٣٦_	٠,٠٧٦٦
٣	٠,٧٢٨٣	٠,٠١٥١	٠,٢٥٦٨	٠,٠٧٣_	٠,١٧٢١	٠,٠٦٤_	٠,١٩٢_
٤	٠,٦٨٨٧	٠,٠٦٥_	٠,١٢٧٩	٠,٠٠٧_	٠,٠٢١	٠,١٦٥٤	٠,٠٠٤٩
١	٠,٦٦٩٢	٠,٠٢٤٩	٠,١١٩٩	٠,١٩٢٨	٠,٠٥٠٧	٠,١٥٦_	٠,٠٨٥
١٩	٠,٥٧٨٧	٠,٠٠٣_	٠,١٨٤_	٠,١٢١٦	٠,١٠٦_	٠,١٩٩٣	٠,٠٧٩_
٧	٠,٠١٢_	٠,٧٩٨٥	٠,١١٧_	٠,٠٣٠_	٠,٠١٨٣	٠,٠٢٤_	٠,١٢٤١
١٠	٠,٠٤٠٩	٠,٧٨٥٦	٠,٠٢٠_	٠,٠٤١٥	٠,٠٣٤_	٠,١٠٣١	٠,٠٦٠٢
١٥	٠,٠٢٧_	٠,٧٦٨٤	٠,٠٥٢	٠,٠٢٢	٠,٠٥٧	٠,٠٥١_	٠,٠٣٧_
٦	٠,٠١٨١	٠,٦٧٩٨	٠,٢١٠٣	٠,٠٤٠٩	٠,١٢٤٥	٠,٠٦١_	٠,١٠١_
١٢	٠,٠٤٣٩	٠,٠٣٢٦	٠,٧٠٩٣	٠,١٠٧١	٠,١٠٤١	٠,٠٧٨_	٠,٠٨٦٧
٥	٠,١٢٨٣	٠,٠٥٧_	٠,٦٧٨٤	٠,٠٥٩٧	٠,٠٦٧١	٠,٠٦٧	٠,٠٠١_
٨	٠,٠٨٨٣	٠,٠١٧	٠,٥٦٠٤	٠,١٢٥٣	٠,١٠٣_	٠,٣٠٩٧	٠,٢٦٥٦
١٤	٠,٠٣٤_	٠,٠٩٥٦	٠,٥٠٣٥	٠,٠٣٧	٠,٠٣١_	٠,٠٧٩٩	٠,٢١٦_
٢٠	٠,١١١٤	٠,٠٠٣_	٠,٢٢٢٦	٠,٧٣٨٧	٠,٠٠٥٢	٠,١١١_	٠,١٦١٩
٢٢	٠,٠٧٦_	٠,٠٢٦٧	٠,٢٠٥٥	٠,٦١٩٦	٠,١٩٨٥	٠,١٨٨٢	٠,٢٣٧_
٢١	٠,٢٧٥٩	٠,١٠٠٥	٠,١٠٨_	٠,٦١٩	٠,١٣٦٢	٠,٠١٨٣	٠,٠٣٢_
١٦	٠,٠٣٤٤	٠,٠٥٢_	٠,٠٤٢_	٠,٠٥٧٤	٠,٨٢٧٧	٠,١١١٦	٠,٠٠١_
٩	٠,٠٤٤_	٠,١٤٠٥	٠,١٠٦٩	٠,٢٢٧١	٠,٧٩٢٢	٠,٠٢٥_	٠,٠٤٢_
١٧	٠,٢٠٤٥	٠,٠٩٤٨	٠,٢٨٤٨	٠,٠٠٠٦	٠,٢٨٦٦	٠,١٩٦٢	٠,١٩٤٦
١٨	٠,١٠٨٢	٠,٦٤٤_	٠,٠٩٠	٠,٠٩٧_	٠,٠٨٨٤	٠,٧٩٣٣	٠,٠٩٧_
٢٣	٠,٠٩٤_	٠,٠٤٧٦	٠,١٦٢٦	٠,٤٤٧٨	٠,٠٧٨١	٠,٥٩٥٤	٠,٠٨٦٢
١١	٠,١٥٨٢	٠,٠٢٩٣	٠,١٥٧	٠,٠٤١٤	٠,١٧٦٢	٠,٠٦٣٣	٠,٧٢٧
٢	٠,١١٥٣	٠,١٠٢٩	٠,٢٩٢	٠,٠٣٢٧	٠,٢٤٩٧	٠,٠٣٠٦	٠,٦٢٣٤

ومن خلال مقارنة النتائج في الجدول (٦) مع التحليل المنطقي للفقرات الاختبارية في بند أدوات الدراسة الفقرة السادسة نستنتج أن العامل الأول يتضمن فقط بعض الفقرات التي تقيس عملية الجمع، وأن العامل الثاني يتضمن فقط بعض الفقرات التي تقيس مفهوم عملية الطرح، وأن العامل الثالث يضمن فقط بعض الفقرات التي تقيس مفهوم عملية الضرب، وأن العامل الرابع يتضمن فقط بعض الفقرات التي تقيس مفهوم عملية القسمة، ونستنتج من ذلك أن العوامل الأربعة

الأولى تقيس فقط مفاهيم العمليات الحسابية الأربع، ويمكن اعتبار ذلك مؤشراً على صدق اختبار الاختيار من متعدد، ولكن تجمع الفقرات المتبقية حول مجموعة أخرى من العوامل قد يضعف من صدق اختبار الاختيار من متعدد.

رابعاً النتائج المرتبطة بالفرضيتين الرابعة والخامسة : التي تهتم بمقارنة معاملات الصعوبة ومعاملات التمييز للاختبارين، ويوضحها الجدول (٧).

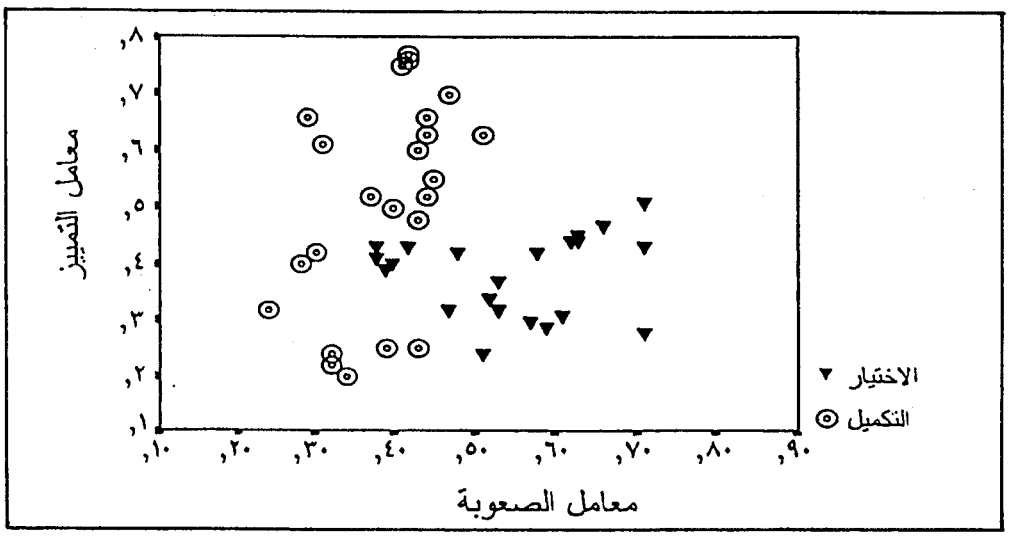
جدول (٧) : معاملات الصعوبة ومعاملات التمييز لفقرات الاختبارين الاختيار من متعدد .

أرقام الفقرات	معاملات الصعوبة		معاملات التمييز ومستوى الدلالة	
	الاختيار من متعدد	التكميل	الاختيار من متعدد	التكميل
١	,٧٠	,٤٧	,٤٧ (٠,٠٠٠١)	,٦٦ (٠,٠٠٠١)
٢	,٥٢	,٣٧	,٣٧ (٠,٠٠٠١)	,٥٣ (٠,٠٠٠١)
٣	,٦٣	,٥١	,٥١ (٠,٠٠٠١)	,٧١ (٠,٠٠٠١)
٤	,٥٥	,٤٥	,٤٥ (٠,٠٠٠١)	,٦٣ (٠,٠٠٠١)
٥	,٦٣	,٤٤	,٤٤ (٠,٠٠٠١)	,٦٣ (٠,٠٠٠١)
٦	,٢٥	,٣٩	,٣٩ (٠,٠٠٠١)	,٣٩ (٠,٠٠٠١)
٧	,٣٢	,٢٤	,٢٤ (٠,٠٠٠١)	,٥١ (٠,٠٠٠١)
٨	,٧٦	,٤٢	,٤٢ (٠,٠٠٠١)	,٥٨ (٠,٠٠٠١)
٩	,٢٥	,٤٣	,٤٣ (٠,٠٠٠١)	,٣٨ (٠,٠٠٠١)
١٠	,٢٠	,٣٤	,٣٤ (٠,٠٠٠١)	,٥٢ (٠,٠٠٠١)
١١	,٤٠	,٢٨	,٢٨ (٠,٠٠٠١)	,٧١ (٠,٠٠٠١)
١٢	,٦٦	,٤٤	,٤٤ (٠,٠٠٠١)	,٦٢ (٠,٠٠٠١)
١٣	,٥٢	,٤٤	,٤٤ (٠,٠٠٠١)	,٦٢ (٠,٠٠٠١)
١٤	,٦١	,٣١	,٣١ (٠,٠٠٠١)	,٦١ (٠,٠٠٠١)
١٥	,٢٤	,٣٢	,٣٢ (٠,٠٠٠١)	,٥٣ (٠,٠٠٠١)
١٦	,٢٢	,٣٢	,٣٢ (٠,٠٠٠١)	,٤٧ (٠,٠٠٠١)
١٧	,٤٨	,٤٣	,٤٣ (٠,٠٠٠١)	,٧١ (٠,٠٠٠١)
١٨	,٦٦	,٢٩	,٢٩ (٠,٠٠٠١)	,٥٩ (٠,٠٠٠١)
١٩	,٤٢	,٣٠	,٣٠ (٠,٠٠٠١)	,٥٧ (٠,٠٠٠١)
٢٠	,٧٧	,٤٢	,٤٢ (٠,٠٠٠١)	,٤٨ (٠,٠٠٠١)
٢١	,٧٥	,٤١	,٤١ (٠,٠٠٠١)	,٣٨ (٠,٠٠٠١)
٢٢	,٦٠	,٤٣	,٤٣ (٠,٠٠٠١)	,٤٢ (٠,٠٠٠١)
٢٣	,٥٠	,٤٠	,٤٠ (٠,٠٠٠١)	,٤٠ (٠,٠٠٠١)

تبين النتائج في الجدول (٧) أن معاملات التمييز لفقرات الاختبارين جميعها دالة إحصائياً على مستوى الدلالة (α < ٠,٠٠٠١) ، وتعد هذه النتيجة دليل صدق للاختبارين، ويبين الجدول أيضاً أن معامل التمييز لكل فقرة من فقرات اختبار التكميل أعلى من معامل التمييز للفقرة المناظرة لها من فقرات اختبار الاختيار من متعدد، عدا الفقرتين (٩) و(٢١) إلا أن الفرق بين معاملي التمييز لهاتين الفقرتين ليس كبيراً. كما يبين الجدول أيضاً أن معامل الصعوبة لكل فقرة من فقرات اختبار التكميل أعلى من معامل صعوبة الفقرة المناظرة لها من فقرات الاختيار من متعدد، وحيث أن معامل الصعوبة يمثل نسبة عدد الطلبة الذين أجابوا على الفقرة إجابة صحيحة إلى عدد الطلبة الذين حاولوا الإجابة عليها؛ فإننا نستنتج من ذلك أن فقرات اختبار الاختيار من متعدد أكثر سهولة من فقرات اختبار التكميل.

كما يبين الشكل (٢) نفس النتائج التي توصلنا إليها في الجدول (٧).

شكل (٢) لوحة انتشار معاملات التمييز ومعاملات الصعوبة لفقرات كل من اختباري التكميل والاختيار من متعدد



كما أظهرت نتائج اختبار (ت) الموضحة في جدول (٨) أن متوسط معاملات التمييز لاختبار التكميل أعلى من متوسط معاملات التمييز لاختبار الاختيار من متعدد ، والفرق بينهما هو فرق دال إحصائياً عند مستوى الدلالة (α < ٠,٠٠٠١) ، أي أن اختبار التكميل أكثر قدرة على التمييز من اختبار الاختيار من متعدد.

جدول (٨) : متوسطات معاملات التمييز والانحراف المعياري للاختبارين وقيمة (ت)

نمط الاختبار	متوسط التمييز	الانحراف المعياري	قيمة (ت)	مستوى الدلالة
التكميل	٠,٥٥٦٧	٠,١١٢	٦,٥٠	٠,٠٠٠١
الاختيار من متعدد	٠,٣٨٨٥	٠,٠٦٨		

بالإضافة إلى ما سبق فقد أظهرت نتائج اختبار (ت) أن متوسط معاملات الصعوبة لاختبار الاختيار من متعدد أكبر من متوسط معاملات الصعوبة لاختبار التكميل، والفرق بينهما هو فرق دال إحصائياً على مستوى الدلالة (٠,٠٠٠١)، أي أن اختبار التكميل أكثر صعوبة من اختبار الاختيار من متعدد، وذلك لأن معامل الصعوبة يمثل نسبة عدد الطلبة الذين أجابوا على الفقرة إجابة صحيحة إلى عدد الطلبة الذين حاولوا الإجابة عليها.

جدول (٩) : متوسط معاملات الصعوبة والانحراف المعياري للاختبارين وقيمة (ت)

نمط الاختبار	متوسط التمييز	الانحراف المعياري	قيمة (ت)	مستوى الدلالة
التكميل	٠,٣٢٩٦	٠,١٥٥	-٩,٢٤	٠,٠٠٠١
الاختيار من متعدد	٠,٥٠٦١	٠,١٨٦		

مناقشة النتائج :

اهتمت الدراسة بإجراء مقارنة بين الخصائص السيكومترية لنمطين من أكثر أنماط الاختبارات الموضوعية شيوعاً هما اختبارات التكميل واختبارات الاختيار من متعدد، إذ تمت المقارنة بين صورتين متكافئتين في المحتوى، حيث سعت الدراسة إلى فحص خمس فرضيات متعلقة بالثبات والصدق والصعوبة والتمييز.

وقد بينت النتائج أن كلاً من اختباري التكميل والاختيار من متعدد يتمتع بدلالات ثبات جيدة، إذ أن معاملات ثباتهما هي معاملات ذات دلالة إحصائية على مستوى الدلالة (α < ٠,٠٠١) ، ولكن معاملي ثبات الإعادة وكرونباخ ألفا لاختبار التكميل أعلى من معاملي ثبات اختبار الاختيار من متعدد بفروق دالة إحصائياً على مستوى الدلالة (α < ٠,٠٠١) ، وهذه النتيجة تتفق مع نتائج دراسة برديجمان التي بينت أن ثبات اختبارات الإجابة المنشأة أعلى من ثبات اختبارات الاختيار من متعدد.

وعلى الرغم من أن هذه النتائج لا تتفق مع آراء كثير من المعلمين وبعض التربويين الذين يرون أن اختبارات الاختيار من متعدد هي الأكثر ثباتاً من أنواع الاختبارات الموضوعية الأخرى، إلا أنها تتفق مع الاتجاهات الحديثة التي ترى أن آراء المعلمين والتربويين حول أفضلية ثبات اختبارات الاختيار من متعدد لا تستند إلى أدلة تجريبية، وأن عامل التخمين الذي لم يستطع العاملون في القياس والتقويم ضبط أثره لا يزال يهدد ثبات اختبارات الاختيار من متعدد.

أما بالنسبة لصدق الاختبار فقد بينت نتائج التحليل العاملي أن الصدق العاملي لاختبار التكميل أفضل من الصدق العاملي لاختبار الاختيار من متعدد، إذ بينت نتائج التحليل العاملي أن عدد العوامل المستخلصة من اختبار التكميل يساوي عدد العوامل التي خطط لقياسها، وأن الفقرات تتجمع حول هذه العوامل بنفس الطريقة التي خطط الباحث لها، في حين زاد عدد العوامل المستخلصة من اختبار الاختيار من متعدد عن عدد العوامل التي خطط لقياسها، أي أن اختبار الاختيار من متعدد يقيس عوامل أخرى غير التي خطط لقياسها، وقد جاءت هذه النتائج متفقة مع نتائج دراسة ثايسن وزملائه التي أظهرت أن اختبار الاختيار من متعدد كان متعدد الأبعاد بالمقارنة مع أبعاد اختبارات الإجابة المنشأة، ولكن بينيت وزملاءه توصلوا إلى أنه يمكن بناء اختبارات الاختيار من متعدد واختبارات الإجابة المنشأة لهما العدد نفسه من العوامل التي تفسر التباين.

ويمكن تفسير اختلاف نتائج الدراسة مع نتائج دراسة بينيت وزملائه إلى اختلاف عدد الفقرات في الاختبارات التي استخدموها في دراستهم، بينما اختير في هذه الدراسة اختباران متكافئان في عدد الفقرات، كما يمكن أن نعزو زيادة عدد

العوامل في اختبار الاختيار من متعدد عن عدد العوامل في اختبار التكميل التي جاءت مطابقة لعدد العوامل المخطط لقياسها، إلى المعرفة الجزئية للمتعلمين ببعض البدائل مما ساعدهم على استبعاد بعضها نتيجة هذه المعرفة الجزئية علماً بأن هذه البدائل قد تقيس أهدافاً أخرى مختلفة عن الأهداف المراد قياسها في تلك الفقرة الاختبارية، الأمر الذي يجعلنا نعتقد بوجود عوامل أخرى مقاسه من خلال هذه البدائل، مما أضعف من صدق اختبار الاختيار من متعدد.

كما بينت النتائج أن معاملات ارتباط فقرات الاختبارين جميعها التي تعد معاملات تمييز للفقرات الاختبارية هي معاملات ذات دلالة إحصائية على مستوى الدلال($\alpha < ٠,٠٠١$)، أي أن فقرات الاختبارين جميعها لها قدرة تمييزية جيدة، وتعد هذه النتيجة دليلاً آخر على صدق الاختبارين، واتفقت هذه النتيجة مع تقديرات المحكمين لمدى ارتباط الفقرات الاختبارية مع الاختبار الذي تنتمي إليه، إلا أن متوسط معاملات التمييز لاختبار التكميل أعلى من متوسط معاملات التمييز لاختبار الاختيار من متعدد، الأمر الذي يجعلنا نعتقد بأفضلية صدق اختبار التكميل على صدق اختبار الاختيار من متعدد، وهو ما يتفق مع نتائج التحليل العاملي، ولكن أحكام المحكمين لم تبين أفضلية صدق اختبار على آخر؛ إذ أن تقديراتهم لمدى ارتباط فقرات اختبار التكميل مع الاختبار نفسه لا تختلف عن تقديراتهم لمدى ارتباط فقرات الاختيار من متعدد مع الاختبار نفسه، وهو ما توصل إليه ثايسن وزملاؤه من أن أسئلة الاختيار من متعدد تقيس ما تقيسه الأسئلة ذات الإجابة المنشأة. وقد يكون من الأسباب التي أدت إلى وجود توافق في أحكام المحكمين أن اهتمامهم أثناء التحكيم لفقرات الاختيار من متعدد نحو نص الفقرة الاختبارية دون التعمق في البدائل، ودون الاهتمام إلى احتمال تأثيرها على ما يمكن أن تقيسه الفقرة، مما حملهم على القول بأن الاختبارين يقيسان الشيء نفسه، على الرغم من نتائج التحليل العاملي مخالفة لذلك.

وقد جاءت نتائج دراسة الصدق منسجمة مع ما يقوله بولاك من أن الأسئلة ذات الإجابة المنشأة أكثر دقة في قياس السمة المراد قياسها من أسئلة الاختيار من متعدد، ومنسجمة أيضاً مع الاتجاهات الحديثة التي يؤيدها كل من هيرمان

وباندي حيث يعتقد أن الاختبارات ذات الإجابة المنشأة أكثر صدقاً من اختبارات الاختيار من متعدد، وذلك لأن ما نختبره من خلال اختبارات الاختيار من متعدد لا يشكل دائماً تناظراً بين ما يقيسه الاختبار فعلاً وبين المعاني الحقيقية أو عمليات التفكير التي نسعى لقياسها، ولأنها قادرة على قياس جميع مستويات الأهداف التي تقيسها اختبارات الاختيار من متعدد، وتقلل من أخطاء التخمين العشوائي، وتقلل من تأثير التغذية الراجعة غير المقصودة، ولكن هذه النتائج جاءت مغاير لتوقعات المعلمين وبعض التربويين غير المدعمة بأدلة تجريبية، الذين يرون أن اختبارات الاختيار من متعدد أكثر أنواع الاختبارات مرونة في قياس مستويات الأهداف المتعددة، وبالتالي أكثر صدقاً من الأنواع الأخرى من الاختبارات الموضوعية.

أما بالنسبة لصعوبة الفقرات الاختبارية في الاختبارين فقد بينت النتائج أن فقرات اختبار التكميل أكثر صعوبة من فقرات اختبار الاختيار من متعدد، وتتفق هذه النتائج مع نتائج دراسة بادجر التي بينت أن أداء المفحوصين على أسئلة الاختيار من متعدد كان أعلى من أدائهم على الأسئلة ذات الإجابة المنشأة. ويمكن تعليل ذلك بأن فقرات اختبار التكميل تتطلب من المفحوص أن يشكل بنفسه البدائل التي سيختار منها الإجابة الصحيحة، حيث يشكل ذلك صعوبة أكبر على المفحوص من اختيار بديل من بين بدائل معطاة، فضلاً عن ذلك فإن وجود البدائل قد سهل على المفحوص الاستفادة من التغذية الراجعة غير المقصودة، التي يحصل عليها إذا كانت الإجابة التي فكر فيها ليست من بين البدائل المعطاة، مما يساعده على اكتشاف خطأ معين، ويحاول تطبيق استراتيجية بديلة لمعرفة الجواب الصحيح، كما أن وجود البدائل قد سهل على المفحوص استخدام استراتيجية الرجوع إلى الخلف التي قد لا تكون هي الهدف بحيث يستخدم لحل السؤال خطوات التأكد من حل السؤال التي قد تكون أكثر سهولة على المفحوص من خطوات حل السؤال نفسه، فمثلا إذا كان السؤال يقيس قسمة عدد على آخر فقد يستخدم المفحوص عملية الضرب في اختيار البديل الصحيح التي تعد أسهل عليه من عملية القسمة.

وفي ختام هذا البحث يوصي الباحث بما يلي:

١) زيادة الاهتمام بأسئلة التكميل في بناء الاختبارات التحصيلية.

٢) مزيد من الدراسات حول الخصائص السيكومترية لفقرات اختبار التكميل ، واختبار الإجابة المنشأة حيث أن الدراسات في هذا الموضوع لا تزال محدودة، وذلك من خلال تطبيق:

٣) نظرية الاستجابة لفقرة (IRT) .

٤ مجموعات دراسية من صفوف أخرى

٥) مهارات رياضية أخرى كحل المسألة.

٦) موضوعات دراسية أخرى غير الرياضيات.

المراجع :

١ – اليونسكو، أبيداس (١٩٨٩). التقرير الختامي للحلقة الدراسية الإقليمية حول تجديد نظم وأساليب التقويم والامتحانات وبنوك الأسئلة في الدول العربية، عمان.

٢ – خليل عليان (١٩٩٤). مناهج البحث العلمي، جامعة القدس المفتوحة، عمان – الأردن.

٣ – ساري سواقد (١٩٩٢). اختبار صحة الافتراضات النظرية لطرق التصحيح لأثر التخمين، ومقارنة أثر استخدام هذه الطرق على الخصائص السيكومترية للفقرة. رسالة دكتوراة غير منشورة. الجامعة الأردنية، عمان ، الأردن.

٤ – عبد الله زيد الكيلاني ، عبد الرحمن عدس (١٩٩٤). القياس والتقويم، جامعة القدس المفتوحة، عمان– الأردن.

٥ – عبد الله زيد الكيلاني (١٩٩٤). التقويم التربوي واختبارات التحصيل، الأنروا/ اليونسكو. دائرة التربية والتعليم، معهد التربية، عمان.

٦ – محمود الخطيب (١٩٨٥). تصنيف أسئلة شهادة الثانوية العامة لمبحث الأحياء ومقارنتها بأسئلة الكتاب المدرسي وأسئلة المعلمين في الأردن، رسالة ماجستير، جامعة اليرموك.

٧ – وزارة التربية والتعليم الأردنية (١٩٩٦). الرياضيات للصف الخامس، الجزآن الأول والثاني، المديرية العامة للمناهج وتقنيات التعليم عمان، الأردن.

٨ – وزارة التربية والتعليم الأردنية (١٩٩٦). الرياضيات للصف الرابع، الجزآن الأول والثاني، المديرية العامة للمناهج وتقنيات التعليم عمان، الأردن.

٩ – وزارة التربية والتعليم الأردنية (١٩٩٦). الرياضيات للصف السادس، الجزآن الأول والثاني، المديرية العامة للمناهج وتقنيات التعليم عمان، الأردن.

10 - Badger, E. (1990). **Using different spectacles to look at student achievement: implications for theory & practice,** from ERIC database computer search ED320938.

11 - Bennett et, al. (1991). Equivalence of free-response and mul tiple-choice items. **Journal of Educational Measurement.** 27(3). Pp. 77-92.

12 - Bridgeman, B. (1992). A Comparison of quantitative Questions in Open-Ended and multiple-choice formats. Journal of **Educational Measurement.** 29(3) 253-271.

13 - Glass, G. & Hopkins, K. (1984). **Statistical Methods in Education and Psychology,** Prentice-Hall, Inc., Englewood Cliffs. New Jersey 07632 PP 305-306.

14 - Gadalla, T. (1999). **Multiple-Choice versus Constructed-Response Tests in the** Assessment of Mathematics Computation Skills, From ERIC database, computer searches ED431813.

15 - Gronlund, N. (1992). **How to Construct Achievement Test.** Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 07632 P 36.

16 - Herman, J. (1992) **What:s Happening with Educational Assessment?** ERIC computer search.

17 - Pandy, T. (1990). **Authentic Mathematics Assessment, computer search,** pp.14,

Available, online: http://erica.net/pare/getvn.asp?v=2&n=1.

18 - Pollack, Judith. (1990). **Some Issues in free Response Testing, computesearch,** from ERIC database, ED322200.

19 - Thissen, D. (et, al.,) (1994). Are Tests comprising both Multiple-choice and free-Response items necessarily less Unidimensional than Multiple-choice Tests? An analysis of two Tests. **Journal of Educational Measurement.** 31 (2) 113-123.

The Effect of the Tests Type on Psychometric Characteristics

Dr. Ahmed E. Al-Kahlout *

ABSTRACT : This study aims at comparing the psychometric characteristics of two types of tests, Multiple Choice and Completions, by examining the following hypotheses:

1) There are no statistical significant differences ($\alpha < 0,05$) between reliability coefficients of the Multiple choice tests and Completion tests.

2) There are no statistical significant differences ($\alpha < 0,05$) between the Judges estimations of the correlation of each item of the Multiple choice test with the test itself, and Judges estimations of the correlation of each item of Completion test with the test itself.

3) There is no difference between the number of significant factors (Eigen value >1), which may be extracted by Factor Analysis technique from the Multiple choice test items and those which may be extracted from the Completion test items.

4) There is no statistical significant difference ($\alpha < 0,05$) between the average of difficulty coefficients of Multiple choice test and average of difficulty coefficients of Completion test.

5) There is no statistical significant difference ($\alpha < 0,05$) between the average of discrimination indices of Multiple choice test and average of discrimination indices of Completion test.

The researcher prepared two forms of tests, Multiple choice and Completion. The completion test items are as the same as items of the Multiple choice test without the alternates. The two forms of tests were applied on a

* Associate Professor, Faculty of Education, AL-Quads open University.

sample of 451 of 6^{th} grade students that were selected from UNRWA schools in Jordan.

The results revealed that the reliability coefficients of the completion test, which were estimated by test-re-test method and Cronbach (α) formula, were better than the reliability coefficients of the multiple choice test. Further the result of (X^2) test revealed that there are no differences between the Judges estimations of the correlation of each item of the Multiple choice test with the test itself, and Judges estimations of the correlation of each item of Completion test with the test itself.

Moreover four significant factors, which are the planned factors, were extracted from the Completion test items, but the number of significant factors, extracted from the Multiple Choice teat items, is seven. Finally the average of difficulty coefficients and the average of discrimination indices of the Completion test items are more than that for the Multiple choice.